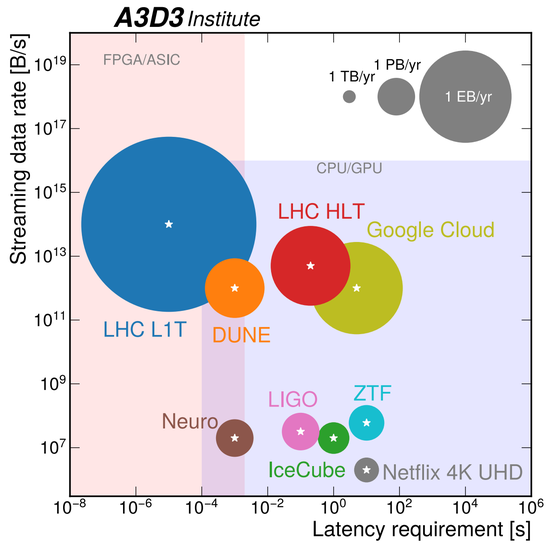
A3D3: Accelerating AI Algorithms

Accelerated AI Algorithms for Data-Driven Discovery

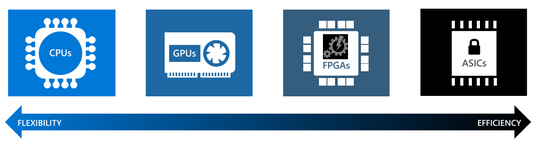
Institutions: *California Institute of Technology, University of California San Diego, Duke University, University of Illinois Urbana-Champaign, Massachusetts Institute of Technology, University of Minnesota Twin Cities, Purdue University, University of Washington Seattle, University of Wisconsin Madison*

A3D3 Institute

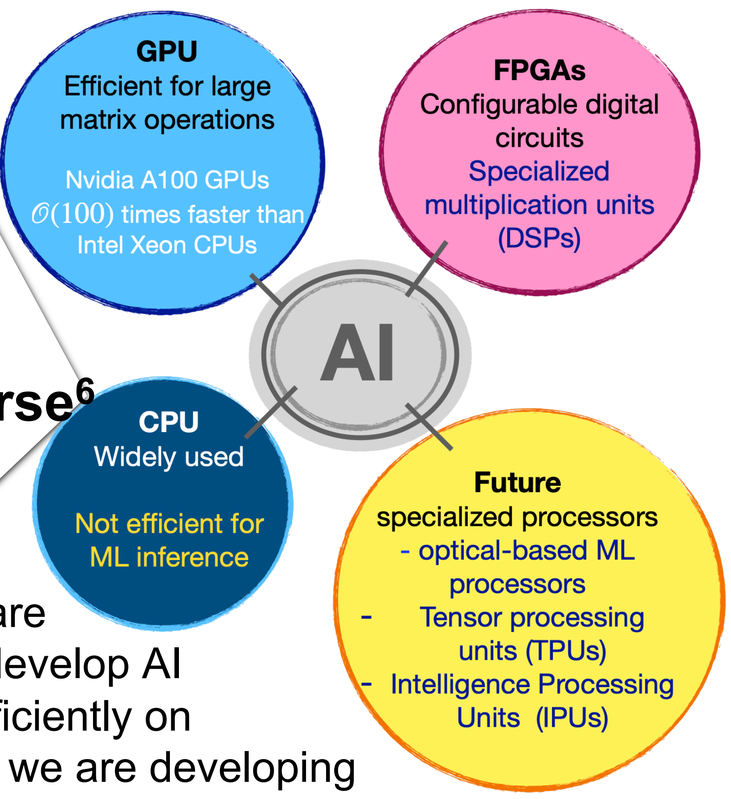
Goals: To pursue next generation AI Algorithms combined with next generation processor technology to develop AI algorithms that can be run *Fast* to solve **real-time scientific problems with AI**
Domains: High Energy Physics, Multi-Messenger Astronomy, Neuroscience



Computational Tools



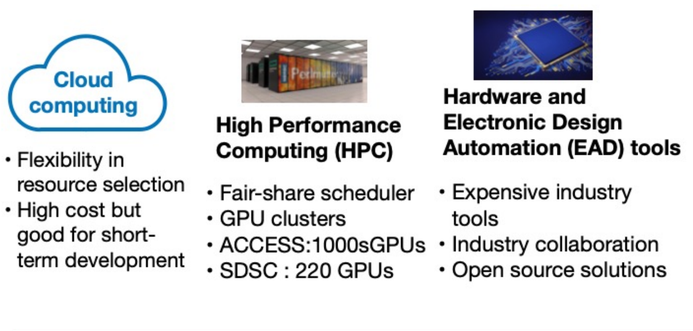
Our Software Tools
ScaleHLS[8], PyLog[7]
SONIC[3],TorchSparse[6]



Computations A major goal of A3D3 is to develop Hardware Aware co-design, where we develop AI Algorithms that can be run efficiently on dedicated hardware. For this, we are developing many tools to run AI on specialized hardware so we can run **Fast ML!**

FPGA Tools To perform low latency batch-1 processing of the information we rely on FPGAs with specialized tools we have developed including HLS4ML,qKeras/qONNX,PyLog,ScaleHLS, and more!

As-a-Service Tools A major goal of our institute is to bring these tools to science as-a-service helps integration, our tools: SONIC,ML4GW,Faast

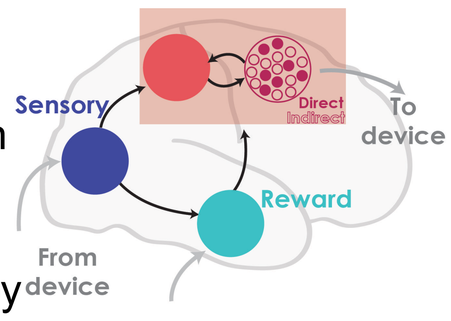


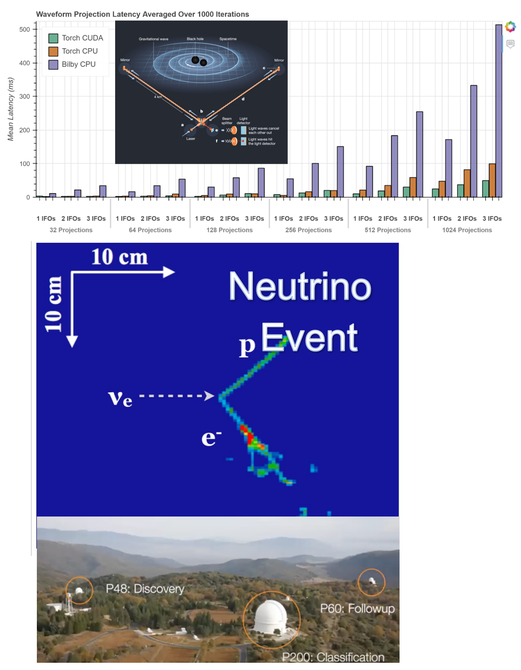
Research Domains

A3D3[1] aims to be a nexus for exchanging new ideas, algorithms and tools between scientific domains, AI communities and industry partners for AI-Hardware co-design

Our focus is on 3 different scientific domains

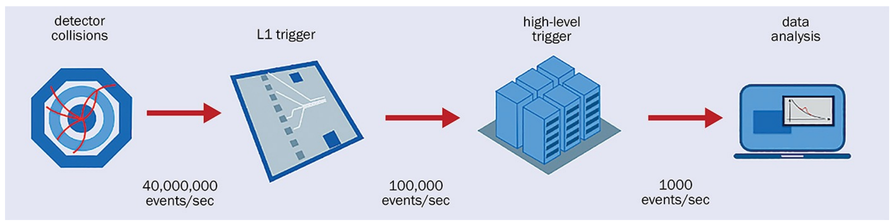
Neuroscience We aim to build AI that can perform real-time readout and control of the Brain through Brain Machine Interfaces. We aim to process behaviors for restorative therapy to gain back functions.



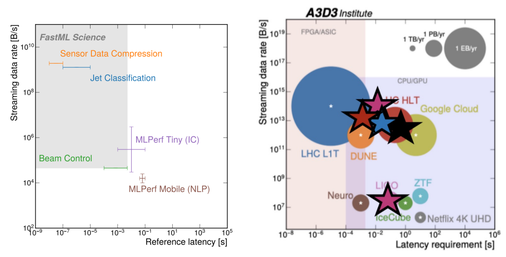


Multi-Messenger Astronomy We aim to build AI that can run real-time AI reconstruction of gravitational waves (LIGO), neutrinos (IceCube, DUNE), and telescope signals (ZTF) rapidly. Sending a signal from one experiment to another (LIGO to telescope) can allow for new profound discoveries.

High Energy Physics We are working on AI algorithms that can process data in sub-microsecond time scales to deal with the ultra low latency challenge needed to process all 40 Million collisions per second with data at over 1 Petabit/second!



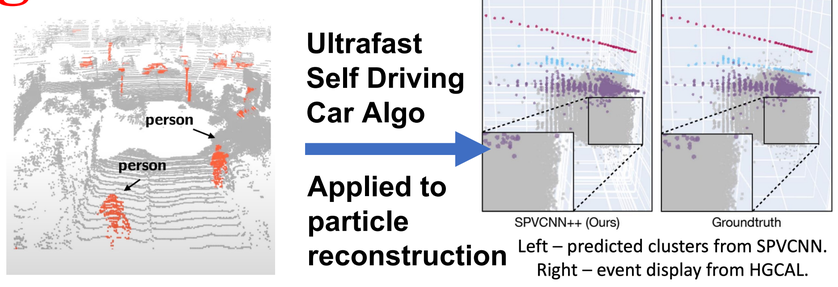
ML Challenges



- Working to make ML challenges to highlight low latency domain[4]
- Highlight our different scientific domains
- Aiming to connect with MLPerf science & Other organizations aimed at scientific challenges

Computing For Science

- Algorithms from our computer science members can be tuned to achieve **optimal low latency and performance for science**. Collaboration for success!



References

[1] https://a3d3.ai
[2] https://fastmachinelearning.org
[3] https://arxiv.org/abs/2007.10359
[4] https://arxiv.org/abs/2207.07958
[5] https://arxiv.org/abs/1804.06913
[6] https://torchsparse.mit.edu/
[7] https://github.com/hst10/pylog
[8] https://github.com/hanchenye/scalehls

This project is supported by the National Science Foundation under Cooperative Agreement" OAC-2117997. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.